数理解析研究所講究録
第 1548 巻 2007 年 91-98

多時点確率取締ゲームの一段階ゲーム戦略

防衛大学校・理工学研究科　前原　裕樹(Hiroki Maehara)
Graduate School of Science and Engineering,
National Defence Academy
防衛大学校・情報工学科　宝崎　隆祐(Ryusuke Hohzaki)
Department of Computer Science,
National Defence Academy

1　はじめに

銃器や薬物の密輸によって不法な資金の獲得を企てる密輸集団と，これの摘発を行うべく活動する取締機関の攻防は日々繰り返されている．ここでは，密輸者と取締機関をプレイヤーとする取締ゲームを考える．従来の研究[1,2]には多時点にわたるモデルが多く，かつ各時点において相手プレイヤーの行動を知ることのできる多段ゲームモデルが大半であった．しかし現実には，相手の採った行動に関する情報を得ることは困難な場合が多い．したがって，相手プレイヤーの戦略に関する情報は得られず，期間中の全ての日の行動を一度に決定する一段階のゲームとして問題を定式化し，その最適戦略を導出する．

2　モデルの前提と定式化

ここでは，パトロールを実施するプレイヤーAと，密輸の実行を企てるプレイヤーBとの間で行われる次のような２人ゼロ和ゲームを考える．

（1）プレイヤーA，Bが１日に１回の行動をとる全体でN日間のゲームを考える．

（2）N日間のなかで，プレイヤーAは最大でK日パトロールを実施可能であり，プレイヤーBは最大でL日密輸が実行可能である．

（3）１回の行動決定に際し，プレイヤーAはパトロールを実施するか否かの２つの手を，プレイヤーBは密輸を実行するか否かの２つの手を持つ．

（4）プレイヤーBの密輸実行日にパトロールを実施することにより，確率pで摘発が成功するが，密輸の成功も確率qで起こる．ただし，$p+q\leqq 1$とし，確率$1-(p+q)$で摘発，密輸がともに生じないとする．また，パトロールが実施されない日に密輸を決行すれば，密輸は必ず成功する．

（5）プレイヤーBが摘発されるか，残り日数が尽きた場合にゲームは終了する．

（6）プレイヤーBの摘発成功によるプレイヤーAの利得は$\alpha(>0)$であり，密輸成功によるプレイヤーBの利得は１である．ただし，パトロール実施日にはプレイヤーBは密輸に出ることはないことを保証するため，$\alpha p-q>0$とする．ゲームの支払をプレイヤーAの利得で定義し，プレイヤーAの利得がプレイヤーBに同量の損失をもたらし，逆もまた真である２人ゼロ和であるとする．

（7）両プレイヤーとも前提（1）～（6）に関し了解しているが，プレイヤーが採った行動は相手プレイヤーには一切知られない．

以上の前提のもとで行われる２人ゼロ和ゲームについて考えていく．まずは幾つかの記号を定義し，各プレイヤーの戦略を表現する．ゲームの行われるＮ日を離散時点$T=\{1,\cdots,N\}$で表現する．時点$i\in T$におけるプレイヤーＡの戦略について，パトロールを実施するならば$x(i)=1$，未実施ならば$x(i)=0$で表す．同様にプレイヤーＢについて密輸実行を$y(i)=1$，未実行を$y(i)=0$で表すと，両プレイヤーの純粋戦略は０及び１の要素を持つＮ次元のベクトル$x=\{x(i),i\in T\}$，$y=\{y(i),i\in T\}$で表される．ただし，前提(2)の最大実行可能回数の制約から

$$\sum_{i=1}^{N}x(i)\leq K\ ,\ \ \sum_{i=1}^{N}y(i)\leq L \tag{1}$$

の制約が課される．

ここで，プレイヤーA，Ｂそれぞれの純粋戦略x，yに対するプレイヤーＡの期待利得を求めると次のようになる．まず，時点ｎでのゲームを考えよう．ゲーム開始時から時点ｎの前日までに両プレイヤーがともに行動を起こす日数を$T(n)$とすると，$T(n)=\sum_{i=1}^{n-1}x(i)y(i)$である．両プレイヤーがともに行動を起こした場合，前提(4)及び(5)から，その日は確率ｐで摘発が起こりゲームが終了することから，前日まで摘発が起こらずに時点ｎに到達する確率は$(1-p)^{T(n)}$である．

また時点ｎでのプレイヤーＡの期待利得について考えると，$(x(n),y(n))=(1,1)$，すなわち両プレイヤーがともに行動を起こした場合は，前提(4)及び(6)からプレイヤーＡは確率ｐでプレイヤーＢを摘発して利得αを得るが，確率ｑで密輸成功を許して１の損失を被るから，期待利得は$\alpha p-q$である．$(x(n),y(n))=(0,1)$，すなわちパトロールが実施されない状況でプレイヤーＢ が密輸を決行した場合は，前提(4)から密輸は確実に成功し，プレイヤーＡは１の損失を被る．$(x(n),y(n))=(1,0)$及び$(0,0)$ の場合，すなわちプレイヤーＢが密輸を実行しない場合には，当然ながら摘発も密輸の成功も起こりえず利得は０である．以上のことから，時点ｎでのプレイヤーＡの期待利得は

$$y(n)x(n)(\alpha p-q)+y(n)(1-x(n))\times(-1)=y(n)\{x(n)(\alpha p-q+1)-1\} \tag{2}$$

と書ける．したがって，全期間における期待支払$R(x,y)$は次式で求められ，これがプレイヤーＡが純粋戦略$x=\{x(i),\ i\in T\}$を，プレイヤーＢが純粋戦略$y=\{y(i),\ i\in T\}$を採った場合のゲームの支払関数となる．

$$R(x,y)=\sum_{n=1}^{N}y(n)\{x(n)(\alpha p-q+1)-1\}(1-p)^{T(n)}. \tag{3}$$

3　支払行列による数値解法

前章における定式化の結果，各プレイヤーは(1)式を満たす有限個の純粋戦略をもち，支払関数は(3)式で与えられることが分かった．したがって，各プレイヤーの純粋戦略を列挙して支払行列を作成し，これに線形計画法を適用することにより問題を解くことができる．しかし，プレイヤーＡの純粋戦略の間には次のような支配関係が存在し，プレイヤーＡは(1)式の制約条件を満たす全ての戦略を使う必要の無いことが分かる．

補題１　プレイヤーＡにとって，パトロール許容回数Ｋを全て行使する戦略は，そうでない戦略を弱く支配す

る.

(証明) プレイヤーAの任意の純粋戦略 $\mathbf{x}$ に対し，ある時点のパトロール戦略をパトロール未実施に置き換えた純粋戦略 $\mathbf{x}'$ は，プレイヤーBの任意の純粋戦略 $\mathbf{y}$ に対し $R(\mathbf{x},\mathbf{y}) \geqq R(\mathbf{x}',\mathbf{y})$ となることから証明される.

ゲームの最適戦略を考える場合，プレイヤーB の純粋戦略としてはN日間で L 回以下の密輸を実行する総数 $\sum_{i=0}^{L} {}_{N}C_{i}$ 通りの戦略を考える必要があるが，プレイヤーAについては補題1から N 日中に K 回のパトロールを実施する総数 ${}_{N}C_{K}$ 通りの純粋戦略を考えればよく，(3)式の支払関数 $R(\mathbf{x},\mathbf{y})$ をもつ行列ゲームとして解くことができる.

4　動的計画法による解法

ここでは，動的計画法を用いて異なった観点からゲームの値について議論していく．前章で述べた線形計画法による均衡解の導出は数値解法であり，これによって得られる数値解からゲームの性質を一般的に議論することは困難である．それを可能にするため，ここでは戦略を変数として取り扱い，解析的にゲームの値を求めることのできる動的計画法による解法を提案する.

4.1　ミニマックス値の導出と密輸者側の最適戦略

ここでは，まずプレイヤーB の任意の混合戦略に対するプレイヤーA の最適な純粋戦略を導出する．これによりプレイヤーA の戦略を最適化することによる期待支払の最大化が行われる．それに引き続いてプレイヤーB の混合戦略を変化させることにより，最大期待支払の最小化を行い，ミニマックス値，すなわちゲームの値を求める.

これまでの議論では，期間中の各日を時間の流れに沿った離散時点で表してきたが，ここでは残り時点数としてステージ番号を定義する．すなわち，時点 $n=1,2,\cdots,N$ をステージ $s=N,N-1,\cdots,1$ で再定義する．これに伴い，時点 $i\in T$ におけるプレイヤーBの戦略表現を，2章の $y(i)$ からステージ t に対して定義される変数 $y_t=y(N-t+1)$ を用いる．因みにプレイヤーB の純粋戦略を $\mathbf{y}=\left\{y_N, y_{N-1},\cdots,y_1\right\}$ で表す．このとき，プレイヤーB の実行可能な純粋戦略の集合は $\mathbf{Y}=\left\{\mathbf{y}\in\{0,1\}^N \,\middle|\, \sum_{t=1}^{N} y_t \leqq L\right\}$ である．さらに，純粋戦略 $\mathbf{y}\in\mathbf{Y}$ を選択する確率を $\pi(\mathbf{y})$ とし，プレイヤーB の混合戦略を $\boldsymbol{\pi}=\left\{\pi(\mathbf{y}),\ \mathbf{y}\in\mathbf{Y}\right\}$ （ただし，$\pi(\mathbf{y})\geqq 0$，$\sum_{\mathbf{y}\in\mathbf{Y}}\pi(\mathbf{y})=1$）で定義する．その他，ステージ t で密輸を実行する純粋戦略の集合を $\mathbf{Z}_t$ と表す．すなわち $\mathbf{Z}_t=\left\{\mathbf{y}\in\mathbf{Y}\mid y_t=1\right\}$ とする．このとき，プレイヤーB が混合戦略 $\boldsymbol{\pi}$ を採用した場合に，ステージ t において密輸が実行される確率は $\sum_{\mathbf{y}\in\mathbf{Z}_t}\pi(\mathbf{y})$ と表される．以下ではプレイヤーB の混合戦略 $\boldsymbol{\pi}$ に対するプレイヤーA の最適な純粋戦略を求めていくが，当面プレイヤーA の戦略の形態を，ステージ t でパトロールを実施する確率を ϕ_t とした行動戦略で議論し，プレイヤーB の混合戦略 $\boldsymbol{\pi}$ に対する最適な行動戦略を求める．結果的には，最適な ϕ_t は0または1となることを予め断っておく.

いま，ステージ t におけるプレイヤーB の混合戦略 $\boldsymbol{\pi}$ が与えられ，残りパトロール可能日数が k である場合

に，ステージｔ以降の最適なパトロール戦略により得られる期待利得の最大値を$f_t^k(\pi)$とする．また，同じ状況にあるステージｔにおいてパトロールを実施するとした場合に，ステージｔ以降で得られる期待利得の最大値を$g_t^k(\pi)$，パトロール未実施とした場合のそれを$h_t^k(\pi)$と定義する．$g_t^k(\pi)$及び$h_t^k(\pi)$はそれぞれ以下の式を満たす．

$$g_t^k(\pi)=(\alpha p-q)\sum_{y\in Z_t}\pi(y)+\left(1-p\sum_{y\in Z_t}\pi(y)\right)f_{t-1}^{k-1}(\Lambda_t\pi), \tag{4}$$

$$h_t^k(\pi)=-\sum_{y\in Z_t}\pi(y)+f_{t-1}^{k}(\pi). \tag{5}$$

ただし，(4)式の$\Lambda_t\pi$は，ステージｔで実施したパトロールによって摘発が起こらなかったという条件の下での混合戦略$\boldsymbol{\pi}$の事後確率を意味し，

$$\Lambda_t\pi(y)=\frac{\pi(y)\left(1-p\,y_t\right)}{1-p\sum_{z\in Z_t}\pi(z)} \tag{6}$$

である．これらの記号を用いると，$f_t^k(\pi)$は次の漸化式で表される．また，ステージｔでの最適なϕ_tの値ϕ_t^*も一緒に書いている．

$$f_t^k(\pi)=\max_{0\le\phi_t\le 1}\left[\phi_t g_t^k(\pi)+(1-\phi_t)h_t^k(\pi)\right]=\max_{0\le\phi_t\le 1}\left[h_t^k(\pi)+\phi_t\left(g_t^k(\pi)-h_t^k(\pi)\right)\right]$$

$$=\begin{cases}(\alpha p-q)\sum_{y\in Z_t}\pi(y)+\left(1-p\sum_{y\in Z_t}\pi(y)\right)f_{t-1}^{k-1}(\Lambda_t\pi)\ , & g_t^k(\pi)>h_t^k(\pi)\text{のとき}\ (\phi_t^*=1)\\ -\sum_{y\in Z_t}\pi(y)+f_{t-1}^{k}(\pi)\ , & g_t^k(\pi)=h_t^k(\pi)\text{のとき}\ (0\le\phi_t^*\le 1)\\ -\sum_{y\in Z_t}\pi(y)+f_{t-1}^{k}(\pi)\ , & g_t^k(\pi)<h_t^k(\pi)\text{のとき}\ (\phi_t^*=0).\end{cases} \tag{7}$$

$$\text{初期条件}：f_0^0(\pi)=0, \tag{8}$$

$$\text{境界条件}：f_t^0(\pi)=-\sum_{i=1}^{t}\sum_{y\in Z_i}\pi(y), \tag{9}$$

$$f_t^t(\pi)=\sum_{y\in Y}\pi(y)\sum_{i=1}^{t}y_i(\alpha p-q)(1-p)^{\sum_{j=i+1}^{t}y_j}. \tag{10}$$

ｔ＝０の場合の(8)式は明らかである．(9)式におけるｋ＝０，すなわち残りパトロール回数が０の場合は，そのステージｔ以降で実行される密輸は確実に成功する．したがって，ステージｉ∈[1,t]において密輸を実行する確率に−１を掛けた期待支払を，ステージ１からｔまで和をとった(9)式のように表される．またｋ＝ｔ，すなわち残りの全ステージにおいてパトロールが実施可能な場合に関しては，補題１からプレイヤーＡは全ステージでパトロールを実施する戦略を採ることとなり，そのときの期待支払が(10)式のように表される．

初期条件から(7)式を用いて逐次計算していくことにより，プレイヤーＢが初期時点のステージＮで混合戦略$\boldsymbol{\pi}$を採った場合の最大期待支払$f_N^K(\pi)$及びプレイヤーＡの最適な戦略ϕ^*が求められる．ただし，(7)式からステージｔにおけるプレイヤーＡの最適な戦略ϕ_t^*は０または１とすればよく，結局はプレイヤーＢの混合戦略$\boldsymbol{\pi}$に対するプレイヤーＡの最適な純粋戦略が求められることになる．

プレイヤーＢの混合戦略$\boldsymbol{\pi}=\{\pi(y),y\in Y\}$は$\pi(y)\ge 0$及び$\sum_{y\in Y}\pi(y)=1$を満たすから，その実行可能領域は$|Y|-1$次元の単位単体を構成する．この領域上で$f_N^K(\pi)$を最小にする点$\boldsymbol{\pi}^*$を見つければ，そこでの

$f_N^K(\pi)$ の値がミニマックス値，すなわちゲームの値であり，π^* がプレイヤーB の最適混合戦略となる．

4.2 密輸者側の最適戦略の導出例

簡単な例として N＝2，K＝1，L＝1の場合に，前節で提案した動的計画法により均衡点を求めてみよう．このときのプレイヤーB の純粋戦略は，２日のうちの何れかの日に密輸を実行する２つの戦略と，密輸を一切行わないという合計３通りの戦略が考えられるから，それぞれの純粋戦略を $y^1=\{1,0\}$，$y^2=\{0,1\}$，$y^3=\{0,0\}$で表すことにする．ただし，時間の流れに沿って要素を並べた表記法$y=\{y(1),y(2)\}=\{y_2,y_1\}$で表現している．このとき，ステージ１，２で密輸を実行する純粋戦略の集合は，それぞれ $Z_1=\{y^2\}$，$Z_2=\{y^1\}$である．純粋戦略 y^1,y^2 を採る確率$\pi(y^1)$，$\pi(y^2)$をそれぞれπ_1,π_2 と簡略化して書くことにすると，y^3を採る確率$\pi(y^3)$は$1-\pi_1-\pi_2$ と表される．したがって，$\sum_{y\in Z_1}\pi(y)=\pi_2$，$\sum_{y\in Z_2}\pi(y)=\pi_1$ となる．(4)～(10)式を用いて逐次計算していくと，プレイヤーB の混合戦略π に対する最大期待支払$f_2^1(\pi)$ は次のようになる．

$$f_2^1(\pi)=\begin{cases}\pi_1(\alpha p-q)-\pi_2, & \pi_1>\pi_2\text{のとき}\quad(\phi_2^*=1,\phi_1^*=0)\\ -\pi_1+\pi_2(\alpha p-q), & \pi_1=\pi_2\text{のとき}\quad(\phi_2^*+\phi_1^*=1)\\ -\pi_1+\pi_2(\alpha p-q), & \pi_1<\pi_2\text{のとき}\quad(\phi_2^*=0,\phi_1^*=1)\ .\end{cases}\tag{11}$$

図１は横軸にπ_1，縦軸にπ_2 をとり，(11)式による$f_2^1(\pi)$ 式の区分を表したものである．また，表１は区分された各領域に対するプレイヤーA の最適戦略$\phi^*=\{\phi_2^*,\phi_1^*\}$ 及び最大期待支払$f_2^1(\pi)$ を表したものである．プレイヤーB の混合戦略の実行可能領域は，$\pi_1\geqq 0$，$\pi_2\geqq 0$，$\pi_1+\pi_2\leqq 1$ を満たす２次元単位単体を構成する．ただし，上述したとおり３番目の純粋戦略y^3を選択する確率π_3 は，$\pi_3=1-\pi_1-\pi_2$ である．上の結果から，

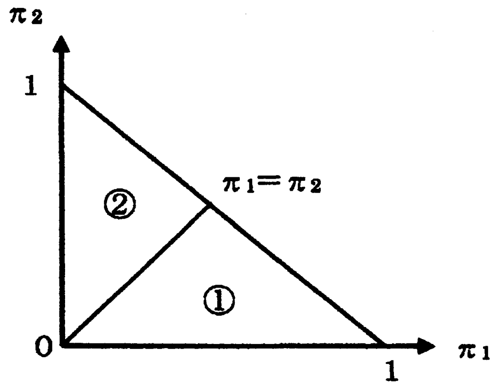

図１：プレイヤーB の戦略の実行可能領域

表１：最適パトロール戦略と最大期待支払

領域	プレイヤーA の最適戦略	$f_2^1(\pi)$
①	$\phi^*=\{1,0\}$	$\pi_1(\alpha p-q)-\pi_2$
②	$\phi^*=\{0,1\}$	$-\pi_1+\pi_2(\alpha p-q)$

プレイヤーAの最適戦略は，$\pi_1=\pi_2$を境界とした２つの領域で異なり，$\pi_1>\pi_2$を満たす領域①内のプレイヤーBの混合戦略に対する最適なパトロール戦略は$\phi^*=\{1,0\}$，最大期待支払は$f_2^1(\pi)=\pi_1(\alpha p-q)-\pi_2$により得られ，領域②内の混合戦略に対しては$\phi^*=\{0,1\}$，最大期待支払は$f_2^1(\pi)=-\pi_1+\pi_2(\alpha p-q)$である．$\pi_1=\pi_2$の境界線上では，パトロールをいずれの日に実施してもよく，$f_2^1(\pi)=\pi_1(\alpha p-q-1)$である．

それでは具体的に$\alpha=2$，$p=0.5$，$q=0.3$と設定して，実際にゲームの値を求めてみよう．このとき$\alpha p-q=0.7$となる．π_1-π_2平面の各点に対し，プレイヤーBの混合戦略に対する期待支払の最大値$f_2^1(\pi)$の大きさを，３次元空間のz軸で表したものが図２であり，$\pi_1=\pi_2$を境界に$z=0.7\pi_1-\pi_2$と$z=-\pi_1+0.7\pi_2$の２つの平面が接する形となる．この最大期待支払を表す平面の最小値，すなわちミニマックス値を与える点$\boldsymbol{\pi}^*$の座標は$(\pi_1,\pi_2)=(0.5,0.5)$であり，その値は-0.15となる．このとき$\pi_3=0$であり，プレイヤーBの最適混合戦略は１回の密輸実行から成る２つの純粋戦略$y^1=\{1,0\}$,$y^2=\{0,1\}$を0.5ずつの確率で採用することであると分かる．次にプレイヤーBにとってより不利な状況を作るため，摘発成功確率pを0.7として，$\alpha p-q=1.1$となる場合を図示したものが図３である．期待支払の増加に伴い，この場合のミニマックス値を与える点$\boldsymbol{\pi}^*$の座標は$(\pi_1,\pi_2)=(0,0)$となり，その値は０となる．このとき$\pi_3=1$であり，プレイヤーBの最適混合戦略は，密輸を行わない純粋戦略$y^3=\{0,0\}$を確率１で採用して利得０を確保することであり，密輸を実行する戦略を採れば期待支払は正となり，プレイヤーA側を利することになる．

以上の２つの具体例から分かるように，最大期待支払を表す平面は，$(\pi_1,\pi_2)=(1,0)$及び$(0,1)$で$z=\alpha p-q>0$，$(\pi_1,\pi_2)=(0,0)$で$z=0$，$(\pi_1,\pi_2)=(0.5,\ 0.5)$で$z=0.5(\alpha p-q-1)$の値をとるから，均衡点は$(\pi_1,\pi_2,\pi_3)=(0.5,0.5,0)$か$(0,0,1)$のどちらかであり，そのどちらになるかは$\alpha p-q$の値に依存する．$\alpha p-q>1$のとき，均衡点は$(\pi_1,\pi_2,\pi_3)=(0,0,1)$でゲームの値は０，$\alpha p-q\leqq 1$ならば均衡点は$(\pi_1,\pi_2,\pi_3)=(0.5,0.5,0)$で，ゲームの値は$0.5(\alpha p-q-1)$となる．

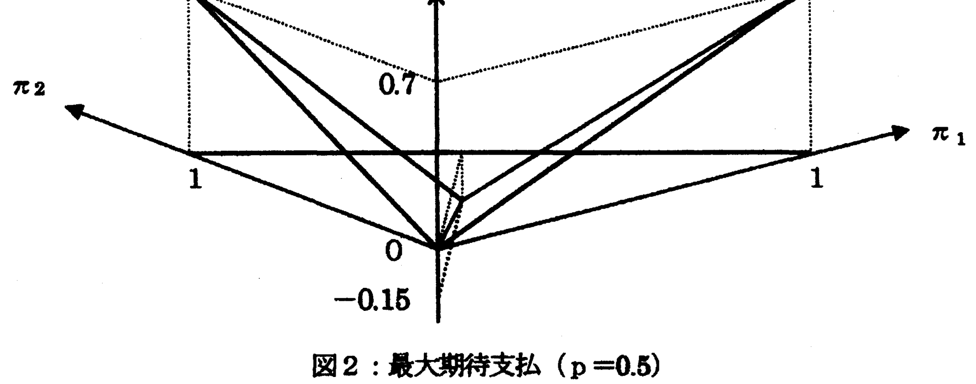

図２：最大期待支払（p＝0.5）

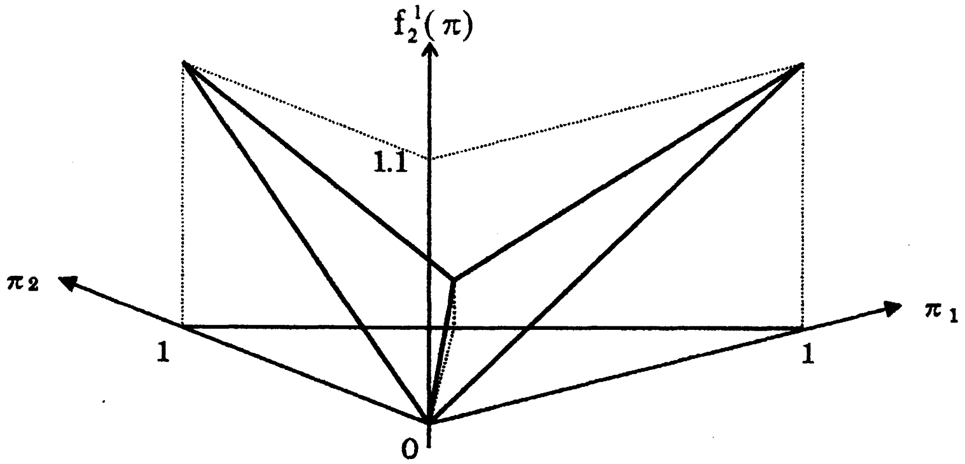

図3：最大期待支払（p＝0.7）

同様に，プレイヤーAの任意の混合戦略に対するプレイヤーBの最適な戦略期待支払の最小化を行い，それに引き続いてプレイヤーAの混合戦略を変化させることにより，最小期待支払の最大化を行い，マックスミニ値を求めることもできる．

5　数値例

線形計画法による数値解法により，N＝7，K＝4，L＝4，$\alpha=1.5$，p＝0.4，q＝0.3 のゲームについて，各プレイヤーの最適戦略を求めた．プレイヤーの最適混合戦略については，プレイヤーA，Bそれぞれの純粋戦略x，yに対する最適な選択確率として得られるが，この性質を見るため，次のような処理を行い，各時点においてパトロールを実施する確率及び密輸を実行する確率を求めた．プレイヤーBの最適混合戦略において，純粋戦略yを選択する確率$\pi^*(y)$により，ステージtで密輸を実行する確率は$\sum_{y\in Z_t}\pi^*(y)$となる．同様にプレイヤーAについても各ステージにおいてパトロールを実施する確率を求め，横軸にステージをとって図示したのが図4である．

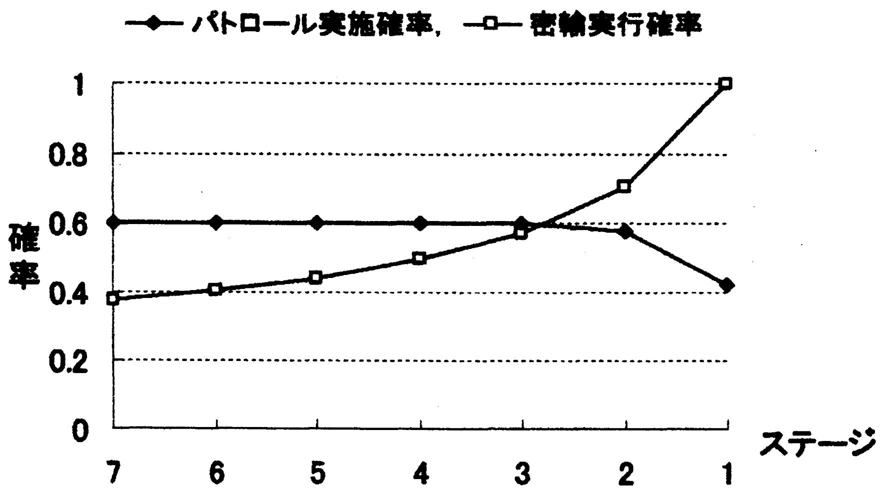

図4：各ステージにおけるパトロール及び密輸実施確率

プレイヤーAの最適戦略は，早期の摘発によってゲームを終了させ，プレイヤーBの横行を阻止することを目指すものである．一方，プレイヤーBは，早い時点で摘発されれば密輸可能回数を残したままゲームを終了することとなるから，早い時点では密輸の実行を控えるのがよいことが分かる．最終日であるステージ1における密輸実行確率は1となっている．このとき，プレイヤーAのステージ1でのパトロール実施確率は0.42である．この場合，プレイヤーBのステージ1での密輸実行による期待利得は，$0.42(\alpha p-q)-(1-0.42)=-0.454$ であり，プレイヤーBにとっては密輸を実行しても期待利益が正となる．なお，このゲームでのゲームの値は-0.62である．

6 おわりに

取締ゲームに関する従来研究では，各時点での相手の行動に関する情報が得られる場合の多段ゲームモデルが主流であったが，本論文では相手プレイヤーの行動が観察できない場合の多時点取締ゲームに対し，線形計画法を用いた一般的な解法に加え，動的計画法による解法を提案した．

ここで取り扱った取締ゲームは，一方のプレイヤーが他方のプレイヤーに同量の損失をもたらす2人ゼロ和ゲームとし，各プレイヤーは定められたパトロール可能回数または密輸実行可能回数内で，各日に行動を起こすかどうかの戦略を採る簡単なモデルであった．

この研究に対する今後の課題として，取締側がパトロールを実施することに伴うコストや，密輸者側が密輸を未実行である場合に課せられるペナルティといった，コスト尺度を取り入れたゲームへの拡張が挙げられる．現実問題として，取締機関にはコストを考慮した効率的な取締活動を行うことが求められており，また密輸者側としては，組織からの指示に従い摘発も覚悟の上で密輸を実行しなければならないケースが考えられる．コストを考慮した場合，これらはプレイヤーの間で異なるのが普通であるため，非ゼロ和ゲームへと拡張しなければならないが，より現実に近い問題になると思われる．

参考文献

[1] M. Sakaguchi, A sequential game of multi-opportunity infiltration, *Mathematica Japonica*, 39, pp.157-166, 1994.

[2] R. Hohzaki, D. Kudoh and T. Komiya, An inspection game taking account of fulfillment probabilities of players, *Naval Research Logistics*, 53, pp.761－771, 2006.